Suppressing sampling noise in linear and two-dimensional spectral simulations

Kruiger, Johannes F.; van der Vegte, Cornelis P.; Jansen, Thomas

Published in:
Journal of Chemical Physics

DOI:
10.1063/1.4907277

Publication date:
2015

Suppressing sampling noise in linear and two-dimensional spectral simulations

Johannes F. Kruiger, Cornelis P. van der Vegte, and Thomas L. C. Jansen

Citation: The Journal of Chemical Physics **142**, 054201 (2015); doi: 10.1063/1.4907277
View online: http://dx.doi.org/10.1063/1.4907277
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/142/5?ver=pdfcov
Published by the AIP Publishing

Suppressing sampling noise in linear and two-dimensional spectral simulations

Johannes F. Kruiger, Cornelis P. van der Vegte, and Thomas L. C. Jansen
Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

(Received 20 October 2014; accepted 20 January 2015; published online 5 February 2015)

We examine the problem of sampling noise encountered in time-domain simulations of linear and two-dimensional spectroscopies. A new adaptive apodization scheme based on physical arguments is devised for suppressing the noise in order to allow reducing the number of used disorder realisations, but introducing only a minimum of spectral aberrations and thus allowing a potential speed-up of these types of simulations. First, the method is demonstrated on an artificial dimer system, where the effect on slope analysis, typically used to study spectral dynamics, is analysed. It is, furthermore, tested on the simulated two-dimensional infrared spectra in the amide I region of the protein lysozyme. The cross polarisation component is investigated, particularly sensitive to sampling noise, because it relies on cancelling of the dominant diagonal spectral contributions. In all these cases, the adaptive apodization scheme is found to give more accurate results than the commonly used lifetime apodization scheme and in most cases better than the gaussian apodization scheme. © *2015 AIP Publishing LLC.* [http://dx.doi.org/10.1063/1.4907277]

I. INTRODUCTION

Two-dimensional optical spectroscopies as two-dimensional infrared (2DIR),[1] sum-frequency generation,[2,3] (2DSFG) two-dimensional visible[4,5] (2Dvis), two-dimensional Raman,[6] VIPER,[7] and fluorescence (2DFS)[8] spectroscopies have recently achieved a lot of attention. These optical techniques were inspired by 2DNMR techniques,[9,10] but typically allow the observation of dynamics on a much shorter time scale.[11–13] The observed resonances are, however, much broader than the corresponding NMR versions resulting in spectra that are much harder to interpret.[14] Therefore, efficient simulation methods are crucial to aid the interpretation.[15–23] While recent theoretical developments allow for the application of simulation methods to 2DIR and 2DSFG of proteins with more than 600 coupled amide I vibrations,[24,25] such calculations are still very time consuming. Therefore, it will be of great importance to find new ways to further improve on the simulation methods. Here, we will examine, how knowledge about the nature of noise arising from limited sampling and the nature of the time-domain response functions can be utilised to suppress the sampling noise. This may lead to speed-up of the calculations without introducing significant spectral aberrations.

The dynamic information in 2D spectra is often analysed by means of slope analysis.[26,27] The spectral slope provides information on the frequency fluctuations induced by the dynamical environment of a chromophore and, thus, reports on this motion. Unfortunately, noise due to insufficient sampling modulates the slope line and leads to inaccuracies of the analysis method.[27,28] Information on couplings can be extracted from 2DIR spectra with the so-called cross polarisation laser arrangement[29–31] that suppresses diagonal peaks and thereby enhances cross-peaks. Simulating this signal requires high accuracy calculations, as the cancellation of the large contributions from the diagonal peaks is not paired with the cancellation of the sampling error. The calculation of 2D IR/vis spectra for complex systems involving hundreds of chromophores is computationally costly despite numerous efforts to extend the commonly used numerical integration of the Schrödinger equation (NISE),[15,20,32–34] the hierarchical equations of motion (HEOM),[23,35,36] classical finite field simulations,[22,37,38] and various non-perturbative approaches.[21,39,40] Reducing the number of disorder realisations needed to get accurate spectra will further help reducing the computation time. All these applications would benefit from a suppression of noise due to limited sampling.

Currently, an artificial lifetime or a so-called apodization function is used to suppress noise in 2D IR/vis experiments and simulations,[14] (see chapter 9) and zero padding may be used to increase the apparent spectral resolution. This type of apodization resembles the finite lifetime of real quantum states,[41] however, this lifetime is often longer than the lifetime needed to get smooth spectra. Therefore, by smoothening, one also obtains spectra that are too broad resulting in overestimation of the homogeneous line broadening component.[14] In this paper, we present a novel method for reducing sampling noise based on knowledge of the physics of the response functions governing the spectroscopic observables.

The remainder of this paper is organised as follows. In Sec. II, we will examine the effect of limited sampling on the simulated spectra. We will discuss the commonly applied lifetime and Gaussian apodization function methods. Using the knowledge of the physical behaviour of the response functions and that of the sampling noise, we will devise a new method for suppressing the effect of limited sampling. In Sec. III, we will apply the new method to two examples. We will pay special

attention to the effect on slope analysis for extracting dynamic information and the cross polarisation arrangement. Finally, we will draw conclusions in Sec. IV.

II. THEORY

To introduce the sampling problem, we will consider the simplest example of a collection of chromophores embedded in an essentially static glassy environment. Each chromophore will then experience a fixed environment dictating the frequency at which that chromophore absorbs light. We will assume that the distribution of these frequencies is Gaussian. The linear absorption spectrum of each individual chromophore will be given by a narrow Lorentzian line, with the width dictated by the lifetime. For a moment, we will assume that this lifetime is very long and can be neglected, so that each chromophore has delta function absorption. Experimentally, one will ideally have a sample with a number of chromophores in the order of Avogadro's number ($\sim 6 \times 10^{23}$). The measured linear absorption spectrum will thus be a smooth Gaussian function given by the underlying heterogeneous distribution of environments (see Fig. 1). For calculating the spectra, we need to sum over a sufficient number of chromophores to get a smooth spectrum (see Fig. 1). The smoothness of the spectrum depends on the chosen resolution and the number of chromophores included in the calculation. The higher resolution and the smoother spectra we want, the more chromophores we need to include.

The sampling problem can also be examined in the time-domain,[42] which is frequently used in practical spectral simulations, in particular, if the environment is not static as in the presently discussed example. The time-domain response function (R) and the linear absorption spectrum are related by a simple Fourier transform.[42] The linear response function of a single chromophore, i, with a fixed frequency, ω_i, is given by

$$R(t) = (\cos(\omega_i t) - i \sin(\omega_i t))\mu_i^2, \tag{1}$$

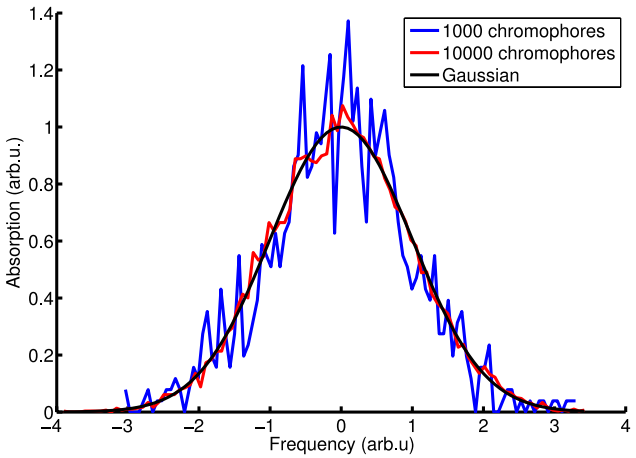

FIG. 1. Black dashed: The spectrum corresponding to a collection of chromophores with a static Gaussian distribution of frequencies. Blue: The spectrum of 1000 chromophores picked randomly from the same Gaussian distribution in a histogram with 100 bins. Red: The spectrum of 10 000 chromophores picked randomly from the same Gaussian distribution in a histogram with 100 bins.

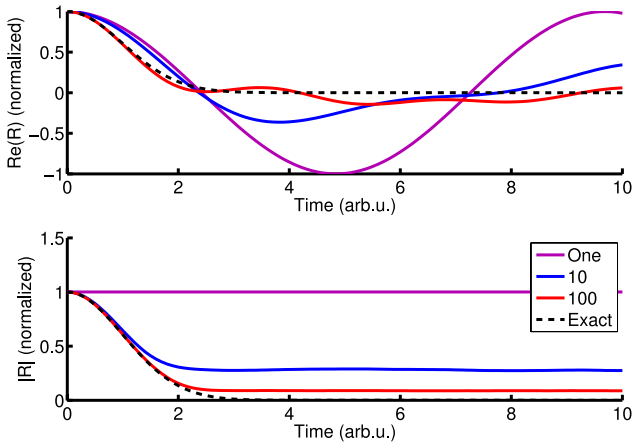

FIG. 2. Top: The real part of the response function for one (green), 10 (blue), 100 (magenta), and exact (black dashed) randomly picked chromophores from the same Gaussian frequency distribution as Fig. 1. Bottom: The absolute value of the response functions for the same number of chromophores as the top panel, but averaged over 1000 realisations of the random selection.

where μ_i is the transition dipole of the chromophore, which will here for simplicity be set to one. We can now look at the structure of the real part of the response function as we average over more and more chromophores in our glassy sample as illustrated in the top of Fig. 2. For a single chromophore, the oscillation will never die out. As we include more disorder realisations, the response function approaches a Gaussian decaying function corresponding to a Gaussian frequency domain spectrum, as expected. However, if we look at the long time behaviour of the response functions, we observe oscillations. These oscillations result in spikes in the spectrum due to the fact that we sampled over a limited number of chromophores. Examining the absolute value response functions, $|R(t)|$, (Fig. 2 bottom) we see that these oscillations disappear with the sampling in a statistical manner resulting in a relative strength of approximately $\frac{1}{\sqrt{N}}$, where N is the number of chromophores included. We also observe that, while the response function that we are interested in starts at one and decays to zero, the sampling noise starts at zero and rises to a constant value. Finally, we observe that the response functions are accurate for short times and they are completely dominated by the sampling noise at later times. The true response of the complete ensemble will decay to zero rather quickly due to dephasing, but the sampling noise will persist forever.

The most frequently used method to eliminate the sampling noise and get smooth spectra is by applying an apodization function.[14] A typical choice is to include an artificial lifetime in the response function simply by multiplying with an exponential decay. This is attractive as, of course, real systems do have a finite lifetime, however, the most optimal lifetime for getting smooth spectra is often shorter than the real lifetime. As the optimal choice of that lifetime is determined by the signal to noise ratio, one could argue that one should simply average over more chromophores. However, this is often not computationally tractable. The effect of the apodization on the spectra is a convolution with the Fourier transform of the apodization function. For an exponential function, we will thus get a convolution with a Lorentzian line shape.[14] Calculating

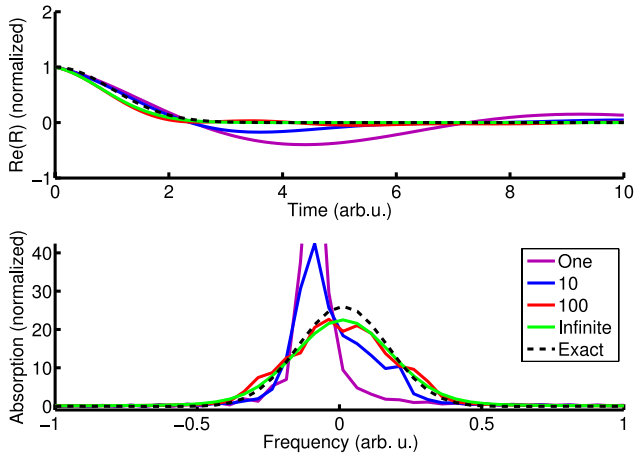

FIG. 3. Top: The real part of the response function including a lifetime of 5 (arb. unit) for the same realisations as in Fig. 2. The exact response function without lifetime is given in dashed black while that with is given in red. Bottom: The corresponding spectra.

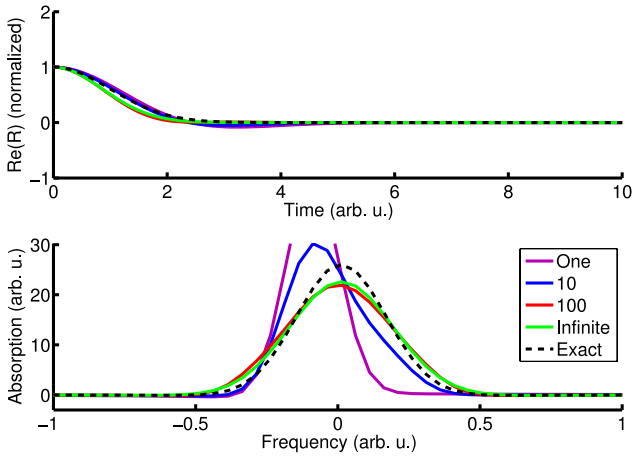

FIG. 4. Top: The real part of the response function including a Gaussian apodization function for the same realisations as in Fig. 2. The exact response function without apodization is given in dashed black, while that with is given in red. Bottom: The corresponding spectra.

the response function for a finite time corresponds to the use of a square apodization function resulting in a convolution of the spectrum with a sinc function. In Fig. 3 (top), the same realisations as in Fig. 2 (top) are shown, but now including lifetime broadening. The first chromophore has a frequency on the red side of the spectrum. Small oscillations are seen in the sideband of the spectrum originating from the convolution with the sinc function. For ten chromophores, the spectrum is improved, but the first chromophore at the red side still makes a significant contribution resulting in a skewed spectrum. For 100 chromophores, the spectrum resembles the final spectrum including a lifetime quite well, but the spectrum is still not really smooth. The spectrum corresponding to an infinite number of chromophores is, of course, very smooth and, as expected, broader than the original one without a lifetime introduced.

One can also employ a Gaussian apodization by multiplying the response function with

$$A(t) = 0.5^{t^2/t_N^2}, \tag{2}$$

where t_N is the transition time between the two regimes. The response functions and spectra using $t_N = 2$ are shown in Fig. 4. The response functions are now very similar, for any number of chromophores. Differences are more visible in the resulting spectra. While all spectra are smooth, the spectra for one and ten chromophores are redshifted compared to the spectrum of the full ensemble. The spectrum for 100 chromophores is hardly discernible from that including the full ensemble and apodization.

Two-dimensional spectra are measured using three incoming laser pulses at different time delays and monitoring the resulting signal emitted from the sample as a function of time. These spectra are governed by third-order response functions.[14] These response functions depend on the three time delays of the interactions between the chromophores and the electric field. The time delays between the two first interactions, t_1, and the time delay between the last two interactions, t_3, are denoted coherence times as the chromophores are in a superposition either between the ground state and a single excited state or in a superposition between a single excited state and a double excited state. The time delay between the second and the third interaction is denoted the waiting time. The coherence times are typically Fourier transformed to obtain the two frequency axes (denoted ω_1 and ω_3). By comparing spectra obtained for different waiting times, dynamical information may be determined from the spectra.[14]

For two-dimensional spectra, apodization functions can be applied as well. However, it has to be generalised to the third-order response. Apodization is applied along both coherence time axes. For the lifetime based version, the function applied is[43,44]

$$\Gamma_{LT}(t_1,t_3) = \exp\left(-(t_1+t_3)/2T_1\right), \tag{3}$$

where T_1 is the (artificial) lifetime. For a Gaussian apodization function, one uses

$$\Gamma_A(t_1,t_3) = \exp(-(t_1^2+t_3^2)/2t_N^2), \tag{4}$$

however, as the signal decay in the diagonal and anti-diagonal direction reflects different physical properties, one can argue that it may be more reasonable to use a function that allows a different transition time in different directions. To accommodate possible difference in the decay of the response function in different directions, one could consider using a function of the form

$$\Gamma_A(t_1,t_3) = \exp(-t_1^2/2T_1^2 - t_3^2/2T_3^2 - t_1t_3/T_C^2), \tag{5}$$

where the different time constants T_1, T_3, and T_C determine the decay of the apodization function in different directions. In general, it is not clear, what the best choice of these constants is and such an approach will, thus, be unattractive.

In the following, a new adaptive apodization method for noise suppression is introduced, first for linear spectroscopy and then for two-dimensional spectroscopy. We first return to Fig. 2, noting that the absolute value of the response function decays to zero and the presence of noise leads to additional response persistent at long times. We now make the anzats that the true response function can be approximated by the

equation

$$R^{\alpha}(t) = R_{\mathrm{sim}}(t)\left(\alpha^{1D}(t)\frac{A_{\mathrm{fit}}(t)}{|R_{\mathrm{sim}}(t)|} + (1-\alpha^{1D}(t))\right). \quad (6)$$

Here, $R^{\mathrm{sim}}(t)$ is the simulated response function, and $A_{\mathrm{fit}}(t)$ is a fit to the absolute value of the simulated response function. $\alpha(t)$ is a scaling function that is zero at short times, when the simulated response can be trusted, and one, when the noise dominates and the correction term should be used. $A_{\mathrm{fit}}(t)$ is determined by fitting $|R_{\mathrm{sim}}(t)|$ to the sum of the functions

$$\begin{aligned} A_{\mathrm{signal}}(t) &= A_S \exp(-t^2/2\tau^2), \\ A_{\mathrm{noise}}(t) &= A_N(1-\exp(-t/\tau)). \end{aligned} \quad (7)$$

The first term has the correct property of the true response starting at some value and decaying to zero as time goes to infinity, while the second term starts at zero and approaches a constant finite value for long times. The three fit parameters correspond to the magnitude of the signal, A_S, the magnitude of the noise, A_N, and a single timescale τ. We now define the function $\alpha^{1D}(t)$ to shift smoothly from zero to one as

$$\alpha^{1D}(t) = \begin{cases} 0 & \text{for } t < t_N \\ (t-t_N)/t_N & \text{for } t_N \le t \le 2t_N, \\ 1 & \text{for } t > 2t_N \end{cases} \quad (8)$$

where t_N is the time, where the fitted signal and the noise are equal in magnitude according to Eq. (7). The result of this adaptive apodization method is demonstrated in Fig. 5. For too few realisations, a negative signal contribution is observed in the low frequency wing. However, for 100 realisations, the resulting spectrum is much closer to the ideal blue spectrum, than for the conventional apodization schemes presented in Figs. 3 and 4.

Now, we will generalise this adaptive apodization scheme to two-dimensional spectra. For the two-dimensional spectra, the absolute value spectra are fitted as a sum of the signal term

$$\begin{aligned} A_{\mathrm{signal}}(t_1,t_3) = A_S \exp(&-(t_1-\tau_1)^2/\tau_2^2 - (t_3-\tau_3)^2/\tau_2^4 \\ &+ (t_1 \times t_3)/\tau_5^2) \end{aligned} \quad (9)$$

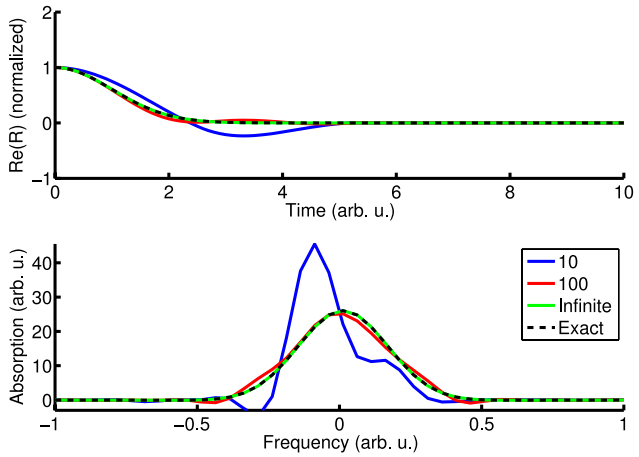

FIG. 5. Top: The real part of the response function including adaptive apodization function for the same realisations as in Fig. 2 except one realisation. The exact response function without apodization is given in red and is identical to the response using an infinite number of realisations and apodization (dashed black). Bottom: The corresponding spectra.

and the noise term

$$A_{\mathrm{noise}}(t_1,t_3) = A_N(1-\exp(-(t_1+t_3)/\tau_6)). \quad (10)$$

As for the linear adaptive apodization noise correction, the response function is corrected using a smooth function, α, now defined in two dimensions as

$$\alpha^{2D}(t_1,t_3) = \begin{cases} 0 & \text{for } \dfrac{A_{\mathrm{noise}}(t_1,t_3)}{(A_{\mathrm{noise}}(t_1,t_3)+A_{\mathrm{signal}}(t_1,t_3))} < A_B \\ \left(\dfrac{A_{\mathrm{noise}}(t_1,t_3)}{(A_{\mathrm{noise}}(t_1,t_3)+A_{\mathrm{signal}}(t_1,t_3))} - A_B\right)/(1-A_B). & \\ & \text{for } \dfrac{A_{\mathrm{noise}}(t_1,t_3)}{(A_{\mathrm{noise}}(t_1,t_3)+A_{\mathrm{signal}}(t_1,t_3))} \ge A_B \end{cases} \quad (11)$$

Resulting in a function that is unity, when the estimated signal to noise ratio is larger than A_B^{-1} and approaching zero, when the estimated signal to noise ratio goes to zero. The constant A_B should be chosen such that α is one, when the signal is expected to dominate, and zero in for coherence times, where the noise is expected to dominate. We then introduce the adaptive apodization correction for the two-dimensional spectra assuming that the simulated response function is given by $R_{\mathrm{sim}}(t_1,t_3)$

$$R^{\alpha}(t_1,t_3) = R_{\mathrm{sim}}(t_1,t_3)\left(\alpha^{2D}(t_1,t_3)\frac{A_{\mathrm{signal}}(t_1,t_3)}{|R_{\mathrm{sim}}(t_1,t_3)|} + (1-\alpha^{2D}(t_1,t_3))\right). \quad (12)$$

This expression, as Eq. (6) for the linear response, ensures that the original simulated response is retained, when the signal to noise is large and suppressed when it is dominated by the noise. The corrected response approaches $R_{\mathrm{sim}}(t_1,t_3)$ as $\alpha^{2D}(t_1,t_3)$ goes to zero. The choice of A_B will be addressed in Sec. III. There we will test the new scheme on more realistic situations, where the true response function deviates considerably from the signal fit functions of Eqs. (7) and (9).

III. RESULTS AND DISCUSSION

To test the adaptive apodization noise correction, the absorption and 2DIR spectra were simulated for a dimer system. A 100 ns long trajectory was constructed for a system with two independent three-level systems. For both vibrations, the frequency fluctuations were constructed to describe an overdamped Brownian oscillator and the standard deviation of the frequency fluctuations was set to 20 cm^{-1}. For the first vibration, the central frequency was set to 1475 cm^{-1}, while the central frequency of the other vibration was set to 1525 cm^{-1}. The frequency correlation times were set to 100 fs and 1 ps for the two vibrations, respectively. The two vibrations were uncorrelated and uncoupled with each other by construction. Both were constructed with an anharmonicity of 20 cm^{-1}. The spectra were simulated using the NISE approach, where the response functions are calculated in the time-domain according to Eq. (11) of Ref. 45. The spectra are then obtained through a two-dimensional Fourier transform over the two coherence times. All spectra were calculated for a zero waiting time.

The linear absorption and two-dimensional spectra were calculated using different intervals for sampling along the full trajectory. This resulted in spectra sampled with different numbers of realisations starting from 10 and ending with 100 000. The original spectra were compared with the spectra using the adaptive apodization noise correction and spectra smoothened using a 1.65 ps lifetime and a 2 ps Gaussian apodization. For the adaptive apodization, the constant A_B was set to 0.05 by maximising the overlap discussed below and defined in Eq. (13) to give the best result for 1000 realisations. The constants used for the lifetime and Gaussian apodization were optimised identically. For the linear absorption, the cutoff time t_N was adjusted to be twice as long as determined by the functions in Eq. (7). This choice was made as the signal from the two peaks interfere with each other resulting in a beating pattern in the absolute value of the response function, thus, reducing the signal size estimated from the fit.

The resulting linear absorption spectra are presented in Fig. 6. The original response is noisier than the ones corrected using the apodization schemes. For the apodization schemes, the spectra are essentially smooth, when using 10 000 realisations, while only for the highest number of realisations, the original spectrum is essentially noise free. For the lifetime apodization, the peaks are broader than in the spectra sampled the best for the other methods. Little difference is seen between the fully sampled spectra of the other apodization schemes. However, for fewer samples, the adaptive apodization gives the smoothest spectra.

The 2DIR spectra are presented in Fig. 7 for four different numbers of realisations. For 100 realisations, the spectra are very noisy, while for 10 000 realisations, the spectra are already almost perfectly smooth, however, the original spectrum is slightly noisier than the corrected ones. A comparison of the 2DIR spectra is performed by looking at the center line slope[26] for the two diagonal peaks and the spectral overlap with the original spectra with the highest number of realisations included in the averaging.

The spectral overlap was calculated as

$$S = \sum_{i,j} \left(I(\omega_i,\omega_j) I_{\mathrm{ref}}(\omega_i,\omega_j)\right) / \sqrt{\left(\sum_{i,j} I(\omega_i,\omega_j) I(\omega_i,\omega_j)\right) \times \left(\sum_{i,j} I_{\mathrm{ref}}(\omega_i,\omega_j) I_{\mathrm{ref}}(\omega_i,\omega_j)\right)}. \quad (13)$$

Here, $I(\omega_i,\omega_j)$ is a two-dimensional spectrum for $\omega_1 = \omega_i$ and $\omega_3 = \omega_j$, while $I_{\mathrm{ref}}(\omega_i,\omega_j)$ is the reference spectrum. This quantity has a value between −1 and 1. When the value is one, the two compared spectra are identical. Values close to one, thus, indicate the best matching with the reference spectrum. Here, we use the original signal with 100 000 realisations as the reference spectrum, as we expect this to be the most accurate. The overlaps between calculated spectra using different numbers of realisations and this reference are presented in Fig. 8. For 100 realisations, the original signal has a overlap below 0.8 with the reference. All apodization schemes give an overlap above 0.9. Looking at the inset of Fig. 8, it is clear that the overlap for all apodization schemes has local maxima for 1000 realisations. This coincides with the point, for which the apodization parameters were optimised. Around 4000 realisations, the original overlap becomes higher than that of all correction schemes, demonstrating the range in which these apodizations schemes are useful. By construction, the adaptive apodization scheme should eventually converge to 1, while the lifetime and Gaussian ones get stuck at constant values slightly above 0.99.

If one is satisfied with an overlap of 0.975 for a given application, one would need about 440 realisations with the adaptive scheme, 550 with the Gaussian, and 700 with the lifetime scheme. Compared to the 1600 realisations needed without apodization, all apodization schemes allow significant reduction in the calculation time, as this time, in general, scales linearly with the number of realisations used. Therefore, the apodization schemes all represent clear computational advantages in this case. For obtaining overlap better than about 0.99, all the presented apodization schemes will fail. One will, thus, need to reoptimize the apodization parameters to obtain an improvement on the original data, when the signal to noise ratio is much higher than what the apodization parameters were optimised for.

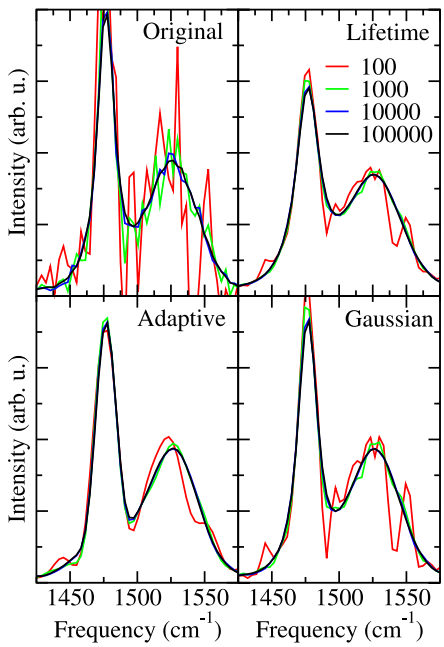

FIG. 6. The linear absorption spectra of the dimer system without apodization (original), lifetime correction (1.65 ps), adaptive apodization, and gaussian apodization (2 ps). The y-axes are identical in all plots.

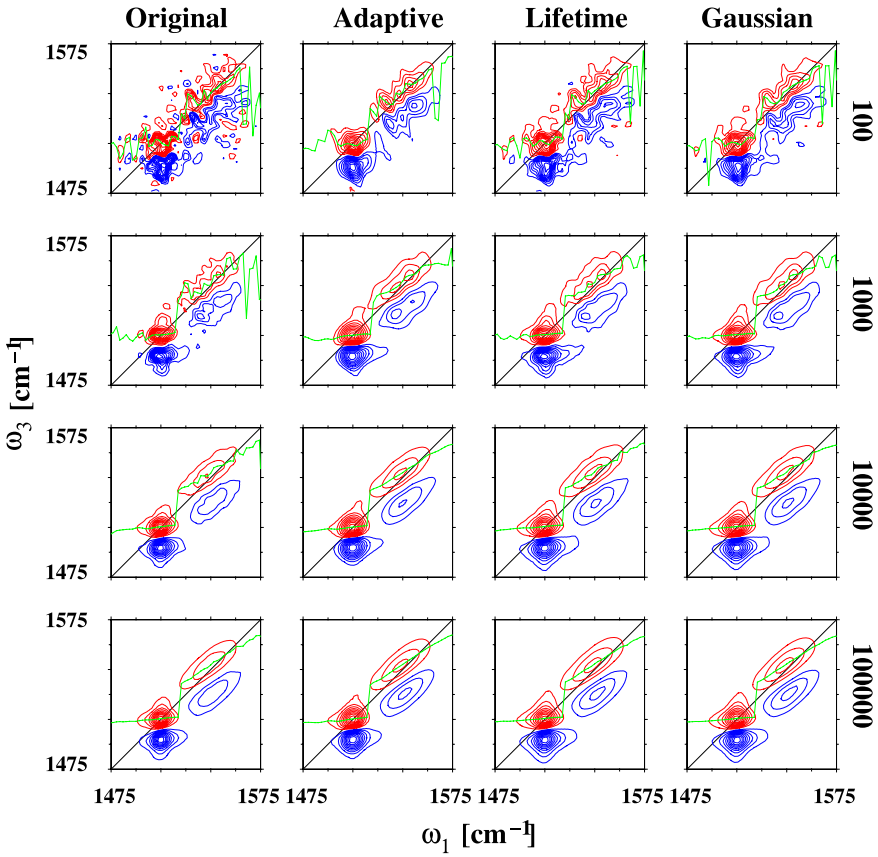

FIG. 7. Two-dimensional spectra for the dimer system. The red contour lines indicate bleach and emission, while the blue contour lines are for absorption. The black line shows the diagonal and the green line is the center line used for slope analysis. The contour lines are drawn equidistant for every 10% of the maximum signal. The rows correspond to different number of realisations used and the columns to the different methods for noise suppression.

The center line slope analysis on the dimer was performed in two regions (1465 to 1485 cm^{-1}) and (1510 to 1545 cm^{-1}) corresponding to the low frequency peak and the high frequency peak. The data are presented in Fig. 9. Large fluctuations in the determined values are found for the original response below 2000 realisations. This is in line with previous studies, where large numbers of realisations were needed to obtain accurate center line slope values,[27,28] which was a major motivation for initiating the present study. The adaptive apodization generally converges faster to the final slope, while the lifetime and Gaussian apodizations systematically give rise to too low slope values. This demonstrates that the adaptive apodization scheme is as efficient as those schemes to suppress noise but does so introducing less undesired spectral aberrations.

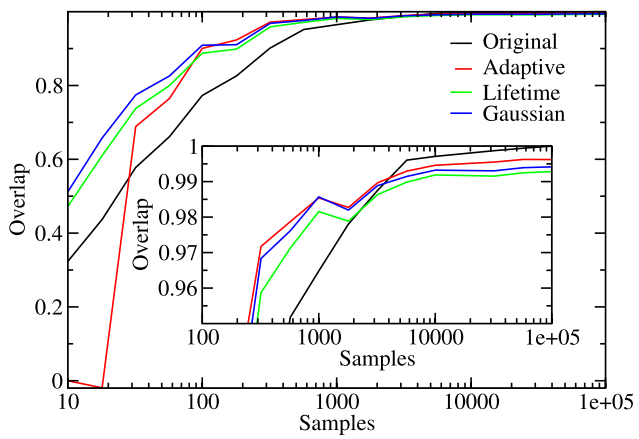

FIG. 8. Peak shape overlap analysis for the dimer using different number of disorder realisations using the original response at the highest sample rate as reference. The inset is a zoom in for the highest number of realisations.

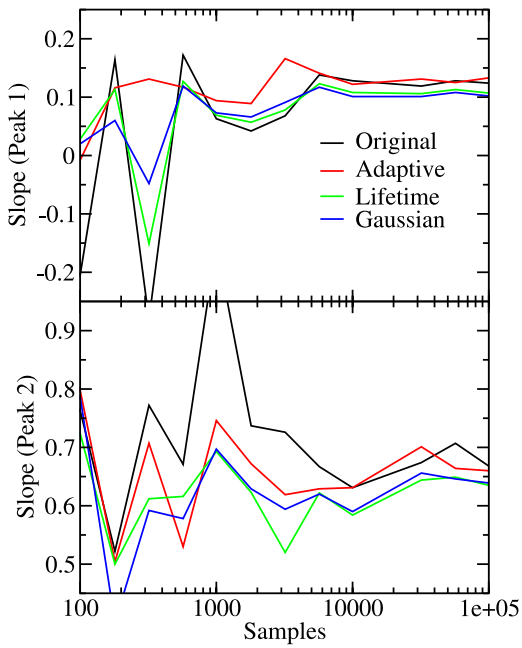

FIG. 9. Slope analysis for the two peaks in the dimer using different number of disorder realisations and apodization schemes.

We now test the adaptive apodization scheme on 2DIR spectra of the amide I band of a protein. For simulating the 2DIR spectrum of the protein lysozyme, a MD simulation was performed using the GROMACS-4.6.5 program[46] with the AMBER 99SB-ILDN[47] and SPC/E[48] force fields using the protein databank structure 1AKI[49] as the starting configuration. After equilibration, a 1 ns production trajectory (NPT) was generated with snapshots stored every 20 fs resulting in a total of 50 000 snapshots. From these snapshots, a vibrational hamiltonian was constructed for the 128 amide I modes in the protein backbone using an electrostatic mapping procedure.[50–53] A fixed site anharmonicity of 16 cm^{-1} (Ref. 1) was employed. The 2DIR spectra were then calculated using the NISE method[15,45,54] as for the dimer example. The 2DIR spectra were calculated by averaging over just 100 different starting configurations along the 1 ns long trajectory corresponding to 100 different disorder realisations. The coherence times, t_1 and t_3, were varied between 0 and 5.12 ps and the waiting time fixed at 0 ps.

The simulated 2DIR spectra for the perpendicular and cross polarization schemes[29,55] are presented in Fig. 10. In particular, the cross polarization choice is challenging to simulate as it involves the cancellation of large signals on the diagonal. In this particular case, the cross polarization signal is about 100 times smaller than the perpendicular polarization signal, which is also clearly reflected in the noise levels observed in the original spectra. Both the adaptive apodization (A_B = 0.05) and the lifetime apodization (T_1 = 1 ps) are much smoother. For the chosen lifetime apodization, commonly used in the simulation of the amide I band of proteins,[24,56,57] the anti-diagonal linewidth is visibly broader than in the original and the adaptive apodization spectra.

The effect of the apodization is finally illustrated by considering the imaginary part of the rephasing and non-rephasing response functions governing the cross polarization 2DIR spectra as shown in Fig. 11. (The real part of the response functions is related to the imaginary part by a Kramers-Kronig like relation and is not shown.[14,42]) In particular, the original non-rephasing response function contains noise at long coherence times, which is effectively suppressed by apodization. The lifetime corrected response decays slightly faster than the adaptive apodization response, in particular, in the diagonal direction.

We demonstrated that the adaptive apodization works well even in general cases, where it is not possible to fit the response functions exactly with the signal fit functions. The difference between the functional form of the fit functions and that of the exact response functions is probably the main reason that the overlap for this scheme did not exceed 0.995, for the dimer case. One could potentially gain accuracy by more flexible fit

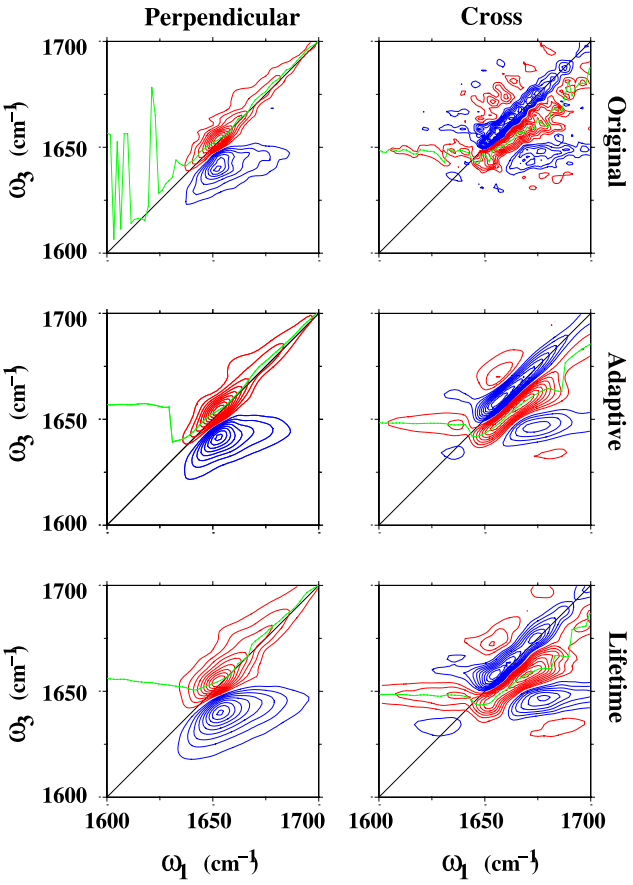

FIG. 10. The 2DIR spectra of the amide I region of lysozyme using the original data, the adaptive apodization (A_B = 0.05), and a lifetime of 1 ps (top to bottom) for the perpendicular and cross polarization setups (left and right). Equidistant contour lines are used for every 10% signal strength with red contours for the bleaching signal and blue for the absorption. Center line slopes are plotted in green.

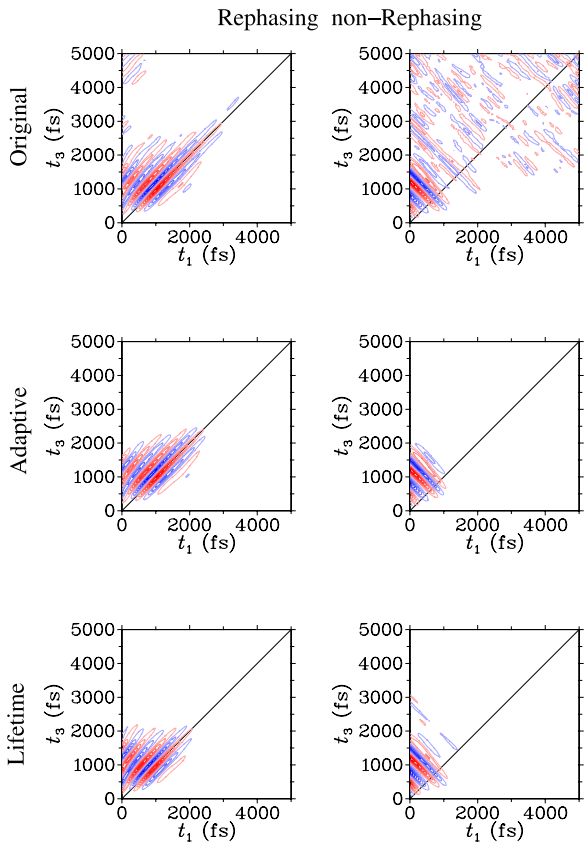

FIG. 11. The rephasing and non-rephasing response function contributions (left and right) for the different apodization schemes. A 1600 cm^{-1} carrier frequency is subtracted to remove fast oscillations in the response functions. Equidistant contour lines are used for every 10% signal strength with red and blue indicating a negative and positive sign, respectively.

functions for the signal. However, using the relatively simple functions presented here, one already achieves improvement as compared to two of the most common apodization schemes. Using more flexible fit functions may also lead to worse results, as the noise may contaminate the fitted signal. If one would, for example, use a biexponential function to fit the signal, one of the terms could get a very long decay time allowing to fit the plateau, which should be described by the growing noise term of the fit. One may, thus, for particular applications be able to improve on the presented scheme by carefully adjusting the functional form used to fit the signal.

In our treatment, we optimised the apodization parameters to get the highest overlap with the reference data for 1000 disorder realisations of the dimer system. Generally, one cannot expect that these parameters are optimal for other applications as well. Unfortunately, in practice one will mostly not know the ideal spectrum needed to optimise these parameters. The adaptive scheme does work well with the same parameters for a broader range of signal to noise ratios than the other apodization schemes investigated. Furthermore, one cannot expect that the optimal apodization parameters found here work, when the functional form of the signal fit function is significantly different from the actual absolute value response function. Future applications on a broader range of systems may provide valuable information on further improving the presented adaptive apodization scheme.

IV. CONCLUSIONS

Here, we demonstrated a new adaptive apodization scheme for suppressing sampling noise in the calculated linear and two-dimensional spectra. It was first demonstrated that the aberrations due to the adaptive apodization are comparable or smaller in linear spectra as compared to more conventional apodization schemes. Using the spectral overlap, the adaptive apodization scheme was found to highly increase the spectral accuracy. It was further demonstrated that the accuracy of center line slope analysis in two-dimensional infrared spectroscopy can be increased, when only a moderate number of realisations can be realised, which is often the case in practical simulations. Finally, the new scheme was demonstrated for spectral simulation of the 2DIR spectrum of Lysozyme. The adaptive apodization was found to be superior for simulating the cross laser polarisation setup, which is used to enhance otherwise spectrally weak cross peaks. Of course, whenever possible simulating more realisations is desirable, however, in practise computational time may be saved by careful use of apodization.

We only considered the effect of the proposed adaptive apodization scheme on 2DIR spectra, however, the new scheme is straight forwardly applicable to other two-dimensional spectroscopies as two-dimensional electronic spectroscopy[4] and two-dimensional sum-frequency generation spectroscopy.[2,58] Furthermore, one could imagine applying the method to experiments, where only few molecules are within the laser spot as in single molecule spectroscopy[59] or two-dimensional nanoscopy.[60] We only demonstrated the method on data obtained with the NISE spectral simulation method, but it can be applied to response functions obtained with other methods as the hierarchical equations of motion, when static disorder is included, or in classical finite field calculations.

ACKNOWLEDGMENTS

Anna Bondarenko is gratefully acknowledged for providing the data for the lysozyme simulations. This work is part of the research programme of the Foundation for Fundamental Research on Matter (FOM), which is part of the Netherlands Organization for Scientific Research (NWO).

[1]P. Hamm, M. H. Lim, and R. M. Hochstrasser, J. Phys. Chem. B **102**, 6123 (1998).
[2]W. Xiong, J. E. Laaser, R. D. Mehlenbacher, and M. T. Zanni, Proc. Natl. Acad. Sci. U. S. A. **108**, 20902 (2011).
[3]J. Bredenbeck, A. Ghosh, M. Smits, and M. Bonn, J. Am. Chem. Soc. **130**, 2152 (2008).
[4]J. D. Hybl, A. W. Albrecht, S. M. G. Faeder, and D. M. Jonas, Chem. Phys. Lett. **297**, 307 (1998).
[5]T. Brixner, J. Stenger, H. M. Vaswani, M. Cho, R. E. Blankenship, and G. R. Fleming, Nature **434**, 625 (2005).
[6]Y. Tanimura and S. Mukamel, J. Chem. Phys. **99**, 9496 (1993).
[7]L. J. G. W. van Wilderen, A. T. Messmer, and J. Bredenbeck, Angew. Chem., Int. Ed. **53**, 2667 (2014).
[8]J. R. Widom, W. Lee, A. Perdomo-Ortiz, D. Rappoport, T. F. Molinski, A. Aspuru-Guzik, and A. H. Marcus, J. Phys. Chem. A **117**, 6171 (2013).
[9]E. L. Hahn, Phys. Rev. **80**, 580 (1950).
[10]C. L. Perrin and T. J. Dwyer, Chem. Rev. **90**, 935 (1990).
[11]S. Woutersen, Y. Mu, G. Stock, and P. Hamm, Chem. Phys. **266**, 137 (2001).
[12]T. L. C. Jansen and J. Knoester, J. Chem. Phys. **127**, 234502 (2007).
[13]J. F. Cahoon, K. R. Sawyer, J. P. Schlegel, and C. B. Harris, Science **319**, 1820 (2008).
[14]P. Hamm and M. T. Zanni, *Concepts and Methods of 2D Infrared Spectroscopy* (Cambridge University Press, Cambridge, 2011).
[15]C. Liang and T. L. C. Jansen, J. Chem. Theory Comput. **8**, 1706 (2012).
[16]J. Manor, P. Mukherjee, Y.-S. Lin, H. Leonov, J. L. Skinner, M. T. Zanni, and I. T. Arkin, Structure **17**, 247 (2009).
[17]A. Paarmann, T. Hayashi, S. Mukamel, and R. J. D. Miller, J. Chem. Phys. **128**, 191103 (2008).
[18]T. L. C. Jansen, D. Cringus, and M. S. Pshenichnikov, J. Phys. Chem. A **113**, 6260 (2009).
[19]J. R. Schmidt, S. A. Corcelli, and J. L. Skinner, J. Chem. Phys. **123**, 044513 (2005).
[20]T. L. C. Jansen and W. M. Ruszel, J. Chem. Phys. **128**, 214501 (2008).
[21]J. Krcmar, M. F. Gelin, and W. Domcke, Chem. Phys. **422**, 53 (2013).
[22]T. Yagasaki and S. Saito, J. Chem. Phys. **128**, 154521 (2008).
[23]L. Chen, R. Zheng, Q. Shi, and Y. J. Yan, J. Chem. Phys. **131**, 094502 (2009).
[24]C. Liang, M. Louhivuori, S. J. Marrink, T. L. C. Jansen, and J. Knoester, J. Phys. Chem. Lett. **4**, 448 (2013).
[25]C. Liang and T. L. C. Jansen, J. Phys. Chem. B **117**, 6937 (2013).
[26]S. T. Roberts, J. J. Loparo, and A. Tokmakoff, J. Chem. Phys. **125**, 084502 (2006).
[27]S. Roy, M. S. Pshenichnikov, and T. L. C. Jansen, J. Phys. Chem. B **115**, 5431 (2011).
[28]A. A. Bakulin, P. A. Pieniazek, J. L. Skinner, T. L. C. Jansen, and M. S. Pshenichnikov, J. Phys. Chem. B **117**, 15545 (2013).
[29]M. T. Zanni, N.-H. Ge, Y. S. Kim, and R. M. Hochstrasser, Proc. Natl. Acad. Sci. U. S. A. **98**, 11265 (2001).
[30]H. Maekawa, C. Toniolo, Q. B. Broxterman, and N. H. Ge, J. Phys. Chem. B **111**, 3222 (2007).
[31]H. Maekawa, F. Formaggio, C. Toniolo, and N. H. Ge, J. Am. Chem. Soc. **130**, 6556 (2008).
[32]M. Yang and J. L. Skinner, J. Chem. Phys. **135**, 154114 (2011).
[33]T. L. C. Jansen, B. M. Auer, M. Yang, and J. L. Skinner, J. Chem. Phys. **132**, 224503 (2010).
[34]C. Falvo, B. Palmieri, and S. Mukamel, J. Chem. Phys. **130**, 184501 (2009).
[35]C. Kreisbeck, T. Kramer, M. Rodriguez, and B. Hein, J. Chem. Theory Comput. **7**, 2166 (2011).
[36]A. Ishizaki and Y. Tanimura, J. Phys. Soc. Jpn. **74**, 3131 (2005).
[37]T. L. C. Jansen, J. G. Snijders, and K. Duppen, J. Chem. Phys. **113**, 307 (2000).
[38]T. Hasegawa and Y. Tanimura, J. Chem. Phys. **128**, 065411 (2008).

[39]P. Kjellberg, B. Brüggeman, and T. Pullerits, Phys. Rev. B **74**, 024303 (2006).
[40]B. Brüggeman, P. Kjellberg, and T. Pullerits, Chem. Phys. Lett. **444**, 192 (2007).
[41]K. Kwac, H. Lee, and M. Cho, J. Chem. Phys. **120**, 1477 (2004).
[42]S. Mukamel, *Principles of Nonlinear Optical Spectroscopy* (Oxford University Press, New York, 1995).
[43]T. L. C. Jansen and J. Knoester, Acc. Chem. Res. **42**, 1405 (2009).
[44]K. Kwac and M. H. Cho, J. Chem. Phys. **119**, 2256 (2003).
[45]T. L. C. Jansen and J. Knoester, J. Phys. Chem. B **110**, 22910 (2006).
[46]H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, Comput. Phys. Commun. **91**, 43 (1995).
[47]K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, R. Dror, and D. Shaw, Proteins **78**, 1950 (2010).
[48]H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, J. Phys. Chem. **91**, 6269 (1987).
[49]P. Artymuik, C. Blake, D. Rice, and K. Wilson, Acta. Chrystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem. **38**, 778 (1982).
[50]T. L. C. Jansen and J. Knoester, J. Chem. Phys. **124**, 044502 (2006).
[51]S. Roy, J. Lessing, G. Meisl, Z. Ganim, A. Tokmakoff, J. Knoester, and T. L. C. Jansen, J. Chem. Phys. **135**, 234507 (2011).
[52]T. L. C. Jansen, A. G. Dijkstra, T. M. Watson, J. D. Hirst, and J. Knoester, J. Chem. Phys. **125**, 044312 (2006).
[53]T. L. C. Jansen, A. G. Dijkstra, T. M. Watson, J. D. Hirst, and J. Knoester, J. Chem. Phys. **136**, 209901 (2012).
[54]H. Torii, J. Phys. Chem. A **110**, 4822 (2006).
[55]R. M. Hochstrasser, Chem. Phys. **266**, 273 (2001).
[56]T. L. C. Jansen and J. Knoester, Biophys. J. **94**, 1818 (2008).
[57]K. Kwac and M. H. Cho, J. Chem. Phys. **119**, 2247 (2003).
[58]J. Bredenbeck, A. Ghosh, H.-K. Nienhuys, and M. Bonn, Acc. Chem. Res. **42**, 1332 (2009).
[59]S. Weiss, Science **283**, 1676 (1999).
[60]M. Aeschlimann, T. Brixner, A. Fischer, C. Kramer, P. Melchior, W. Pfeiffer, C. Schneider, C. Strüber, P. Tuchscherer, and D. V. Voronine, Science **333**, 1723 (2011).